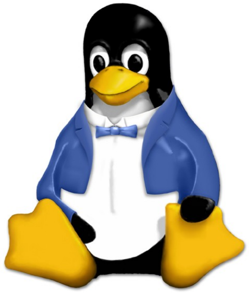


Open Source High Availability on Linux

Alan Robertson
alanr@unix.sh
OR alanr@us.ibm.com

HighAvailability

October 2006

© 2006 IBM Corporation

Agenda - High Availability on Linux

- HA Basics
- Open Source High-Availability Software for Linux
  - Linux-HA Open Source project
  - DRBD Open Source Project
  - Linux Virtual Server (LVS) Project

The Desire for HA Systems

Who wants low-availability systems?

Why are so few systems High-Availability?

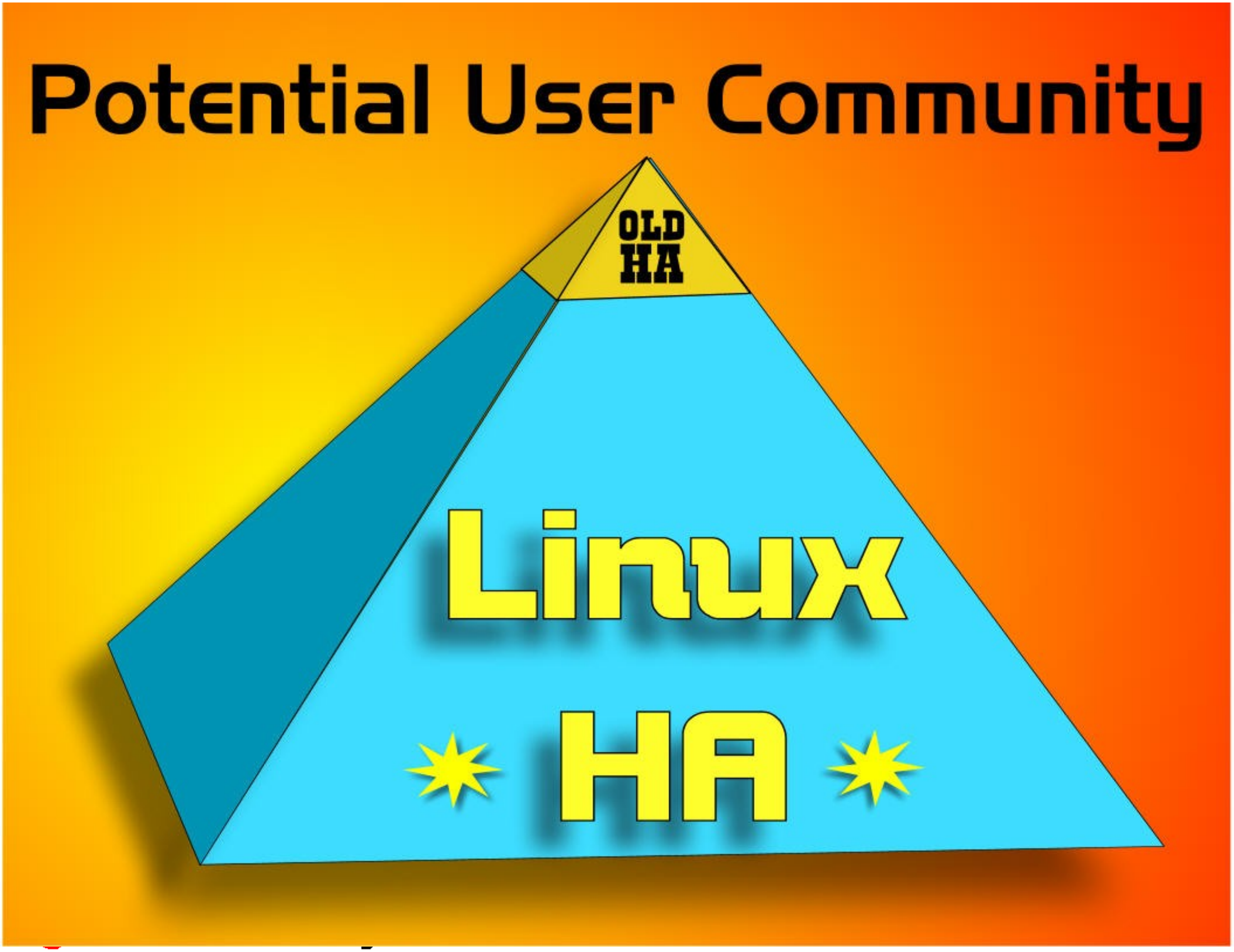
HighAvailability

© 2006 IBM Corporation

Barriers to HA Systems

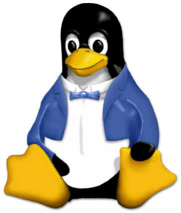
- Cost
  - Very manageable with modern hardware, OSS software
- Complexity
  - Can't give away 'simplicity' – good management tools help

Potential User Community



What would be the result?

- Increased Availability
- Drastically multiplying customers multiplies experience - products mature faster (especially in OSS model)
- OSS developers grow from customers
- **OSS Clustering is a disruptive technology**

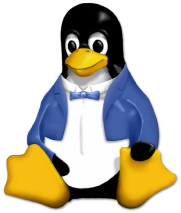


What is a Computer Cluster?

- From Wikipedia:

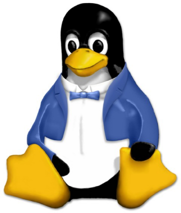
  A computer cluster is a group of loosely coupled computers that work together closely so that in many respects they can be viewed as though they are a single computer.

  Clusters are usually deployed to improve performance and/or availability over that provided by a single computer, while typically being much more cost-effective than single computers of comparable speed or availability.



HA vs. HPC Clustering

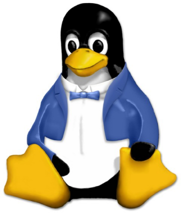
- HPC clusters work primarily to manage and maximize the increased performance which results from having multiple computers working together
- High-Availability clusters primarily work to manage and maximize the increased availability which is possible when multiple computers work together
- These goals are not mutually exclusive



© 2006 IBM Corporation

What is an HA cluster?

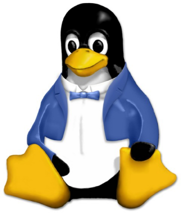
- A group of computers which cooperate to provide a service even when system components fail
- When one machine goes down, others take over its work
  - This involves IP address takeover, service takeover, etc.
  - New work comes to the “takeover” machine
- When a service fails, it is restarted
  - Can be restarted on the same server or a different one



]

What Can HA clustering do for you?

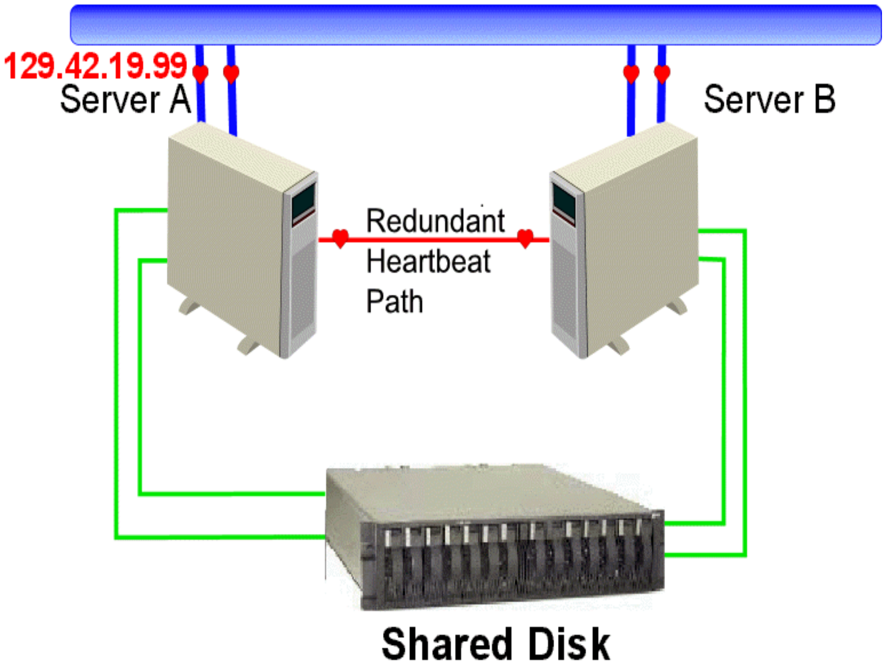
- **It cannot achieve 100% availability** – *nothing can.*
- HA Clustering primarily designed to recover from single faults
- It can make your outages very short
  - From about a second to a few minutes
- It is like a Magician's (Illusionist's) trick:
  - When it goes well, the hand is faster than the eye
  - When it goes not-so-well, it can be reasonably visible
- A good HA clustering system adds a “9” to your base availability
  - 99->99.9, 99.9->99.99, 99.99->99.999, etc.
- **Complexity is the enemy of reliability!**



© 2006 IBM Corporation

High Availability Approach - Redundancy

- Redundancy eliminates Single Points Of Failure (**SPOF**)
- Reduces cost of planned and unplanned outages



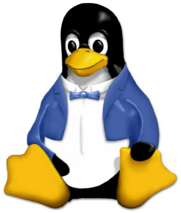
HighAvailability

© 2006 IBM Corporation

The 3 R's of High-Availability

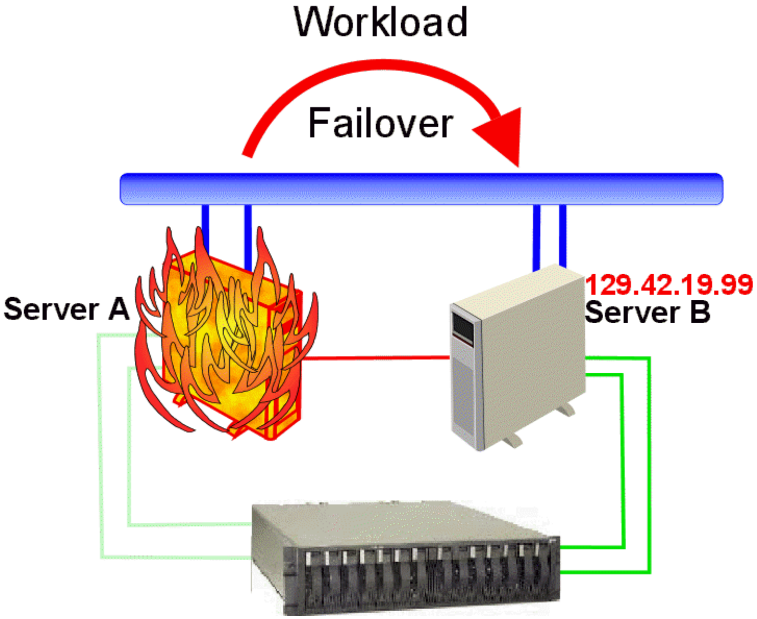
- **R**edundancy
- **R**edundancy
- **R**edundancy

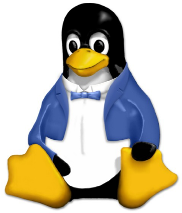
- If this sounds redundant, that's probably appropriate...
  ;-)
- **HA Clustering is a good way of providing and managing redundancy**



High Availability Approach - Failover

- Auto detect Failures (hardware, network, applications)
- Automatic Recovery from failures (no human intervention)
- Managed failover to standby syste components



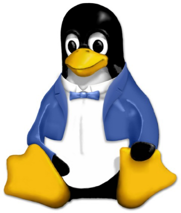


© 2006 IBM Corporation

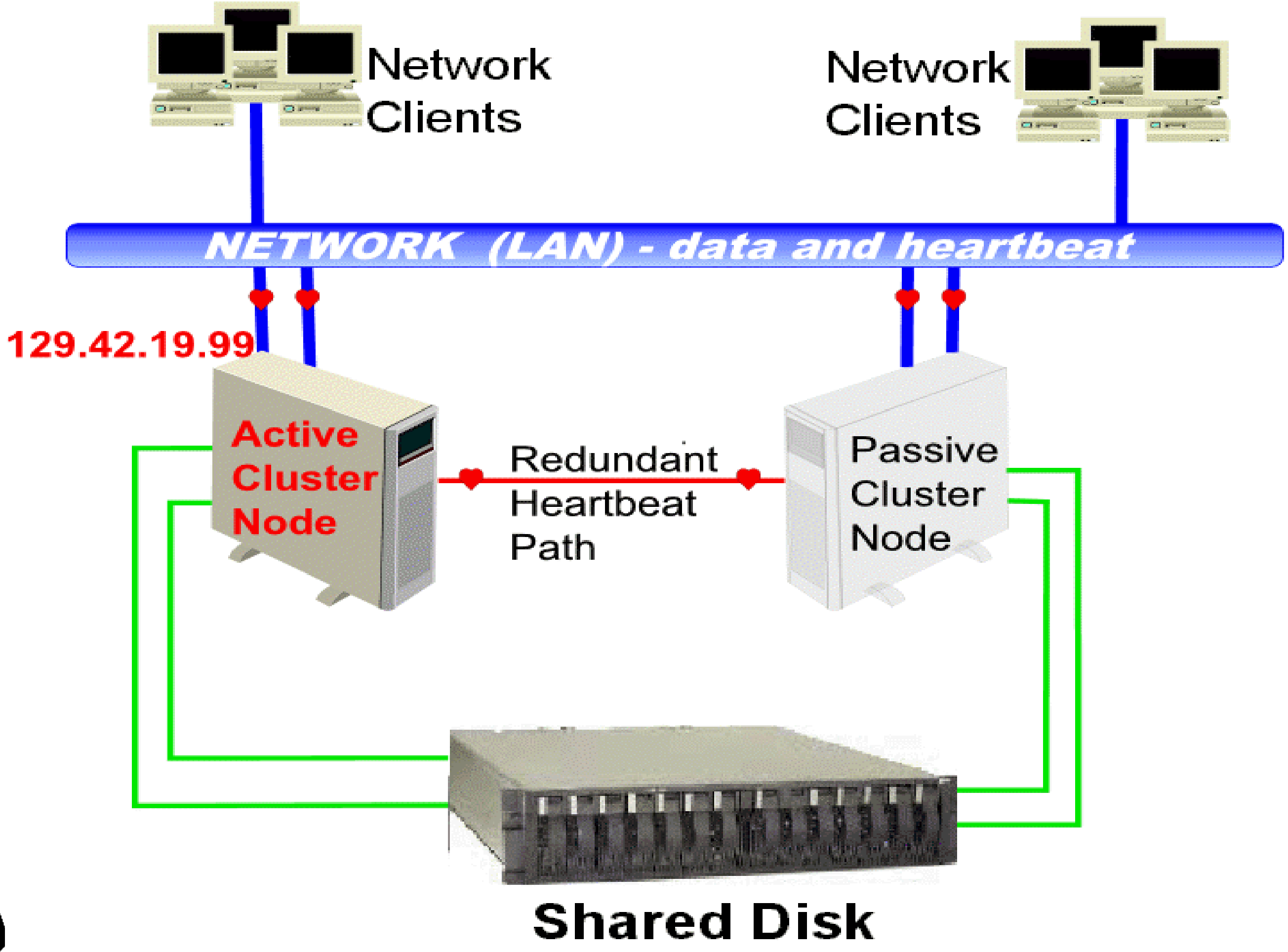
Statistics... Counting Nines...

Availability percentage	Yearly downtime
100%	0
99.99999%	3s
99.9999%	30 sec
99.999%	5 min
99.99%	52 min
99.9%	9 hr
99%	3.5 day

HighAvailability

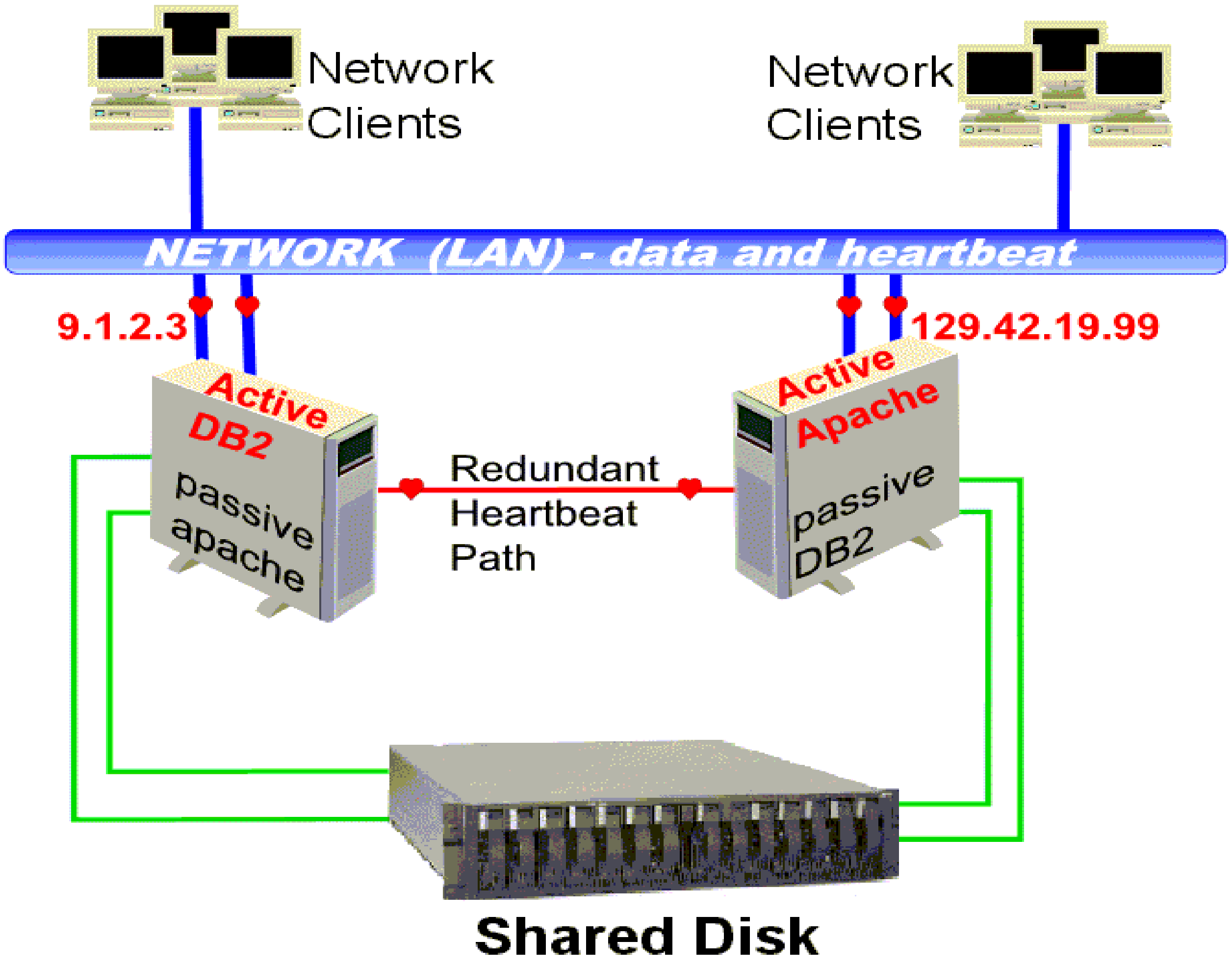


Two Node Active/Passive HA Cluster
Shared Disk (DS4000, ESS, etc.)



© 2006 IBM Corporation

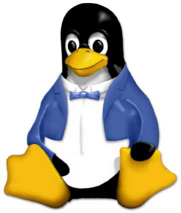
Two Node Active/Active HA Cluster Shared Disk (DS4000, ESS, etc.)



© 2006 IBM Corporation

Linux-HA ("heartbeat") Project

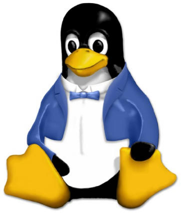
- Open Source Project (IBM Leadership)
- Multiple platform solution for Linux, Solaris, *BSD, OS/X
- Packaged with most Linux Distributions (except Red Hat)
- Part of OSCAR-HA package
- Strong focus on ease-of-use, security, low-cost
- > 30K clusters in production since 1999
- Equal to or superior to commercial HA packages



What is the "Linux-HA" project?

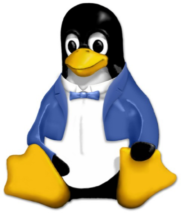
- An open-community project providing basic fail over capabilities for Linux (and other OSes)
- Active, open development community led by IBM
- Wide variety of industries, applications
- Reference implementation for Open Cluster Framework (OCF) standards
- Simple to understand and easy to install
- No special hardware requirements; no kernel dependencies, all user space
- All releases tested by automatic test suites
- http://linux-ha**.org/**

HighAvailability



"Linux-HA" Successes

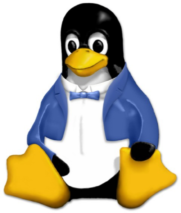
- FexEx – used in truck scheduling
- The Weather Channel (weather.com)
- BBC – internet infrastructure
- CERN – grid services
- Los Alamos National Laboratories – badge readers
- Sony - manufacturing processes
- United Nations
- Intuit (Quicken, TurboTax, etc.) use it for firewalls
- Agilent Technologies in Fort Collins – 3 clusters
- ISO New England manages the New England power grid using 12 "Linux HA" clusters
- University of Toledo – 20K user WebCT System
- Emageon – medical imaging services
- ADC – telco provisioning manager product (w/ x330/335)
- Incredimail uses "Linux HA" on IBM hardware
- Bavarian Radio Station (Munich) used "Linux HA" and xSeries for coverage of 2002 Olympics in Salt Lake City
- More listed at: http://linux-ha.org/SuccessStories



Linux-HA Capabilities

- Supports n-node clusters – where 'n' is currently <= something like 16
- Active/Passive or full Active/Active
- Can use UDP bcast, mcast, ucast comm.
- Fails over on node failure, or on service (resource) failure
- Fails over on loss of IP connectivity, or arbitrary criteria
- Support for the OCF resource management standard
- Sophisticated dependency model with rich constraint support (resources, groups, incarnations, master/slave)
- XML-based resource configuration
- Configuration and monitoring GUI
- Support for OCFS2 cluster filesystem – others coming

HighAvailability

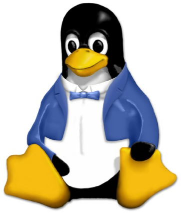


Linux-HA futures being considered

- Business Continuity support (in source control now)
- Specific virtualization support
  - Transparent migration
  - “Containerized” resources (peek inside client VM via proxy)
- Increase number of nodes directly supported
- Loosen cluster definition to manage *many* more nodes through hierarchical proxies
- Integration with provisioning software

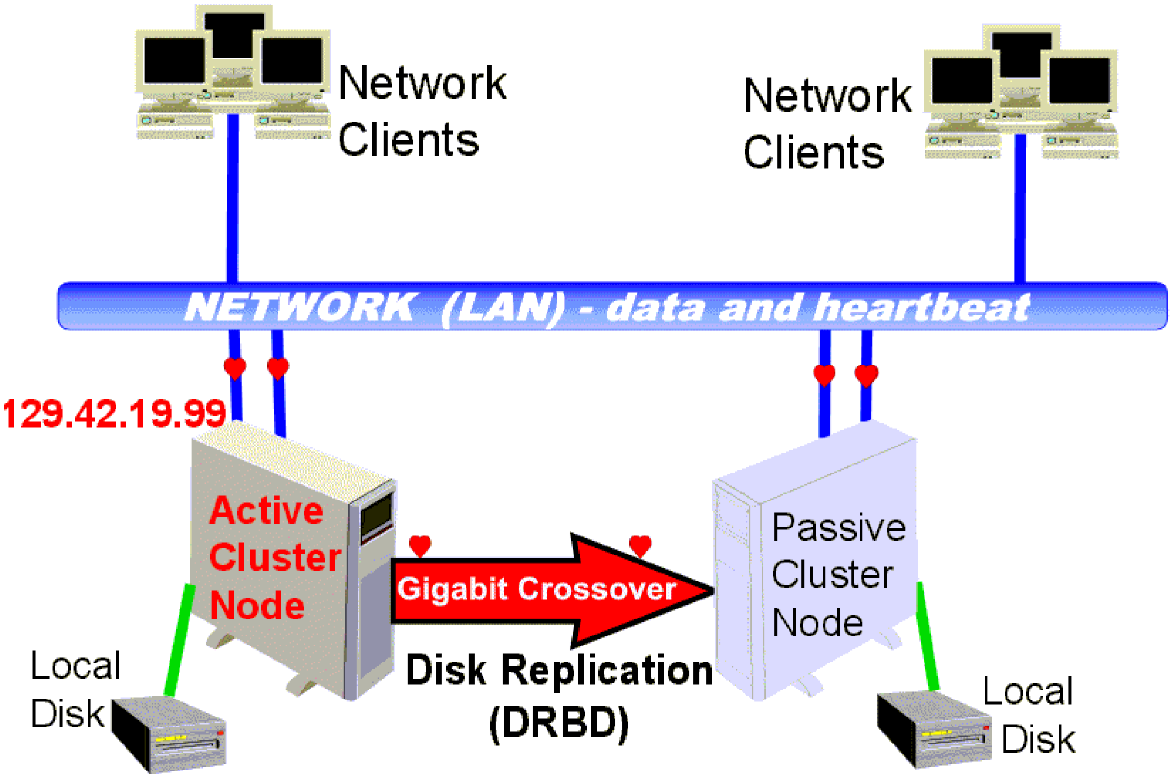
DRBD – Distributed Replicating Block Device
RAID1 over the LAN

- DRBD is a block-level replication technology – it works underneath any (non-clustered) filesystem
- Every time a block is written on the master side, it is copied over the LAN and written on the slave side
- It is *extremely* cost-effective – common with xSeries
- Typically, a dedicated replication link is used
- Also used with slower links for Business Continuity
- Worst-case around 10% throughput loss – typically negligible
- Current versions have very fast "full" resync



© 2006 IBM Corporation

Two Node Active/Passive HA Cluster
Real-Time Disk Replication (DRBD)
DRBD = Distributed Replicating Block Device

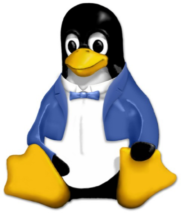


HighAvailability

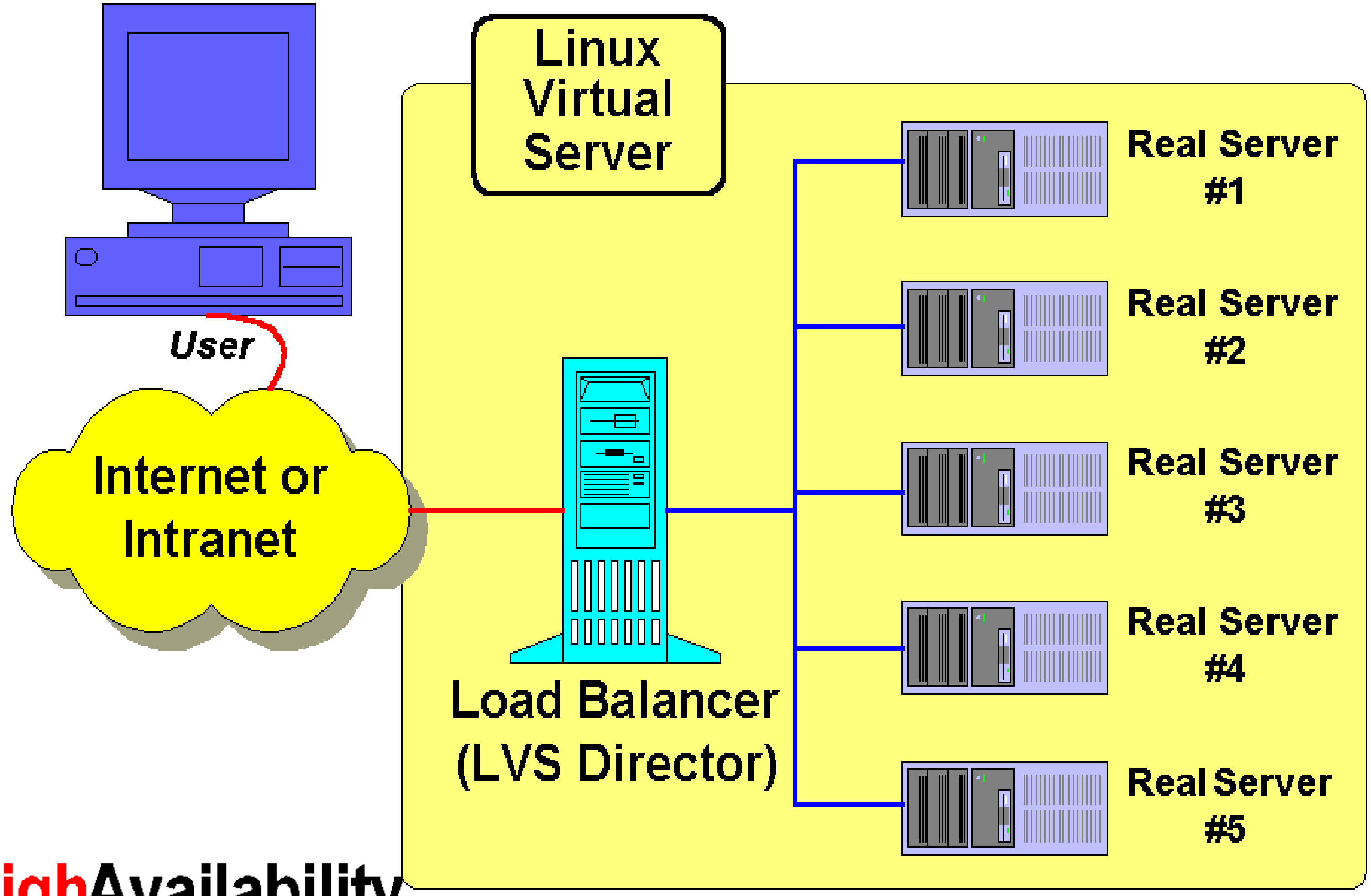
© 2006 IBM Corporation

Linux Virtual Server (LVS) Project

- Linux Virtual Server (LVS/ipvs) comes with Linux, very widely used
  - IP sprayer type of load balancer
  - Commonly used in “server farm” type arrangements
  - Integrates well with Linux-HA
  - Used in many mission-critical applications (like medical imaging, credit card authorization, nuclear facilities)
  - Some customers perform stateful load-balancer failover in less than .5 seconds
  - Support for stateful active/active load balancer clusters



LVS In Action

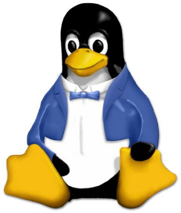


HighAvailability

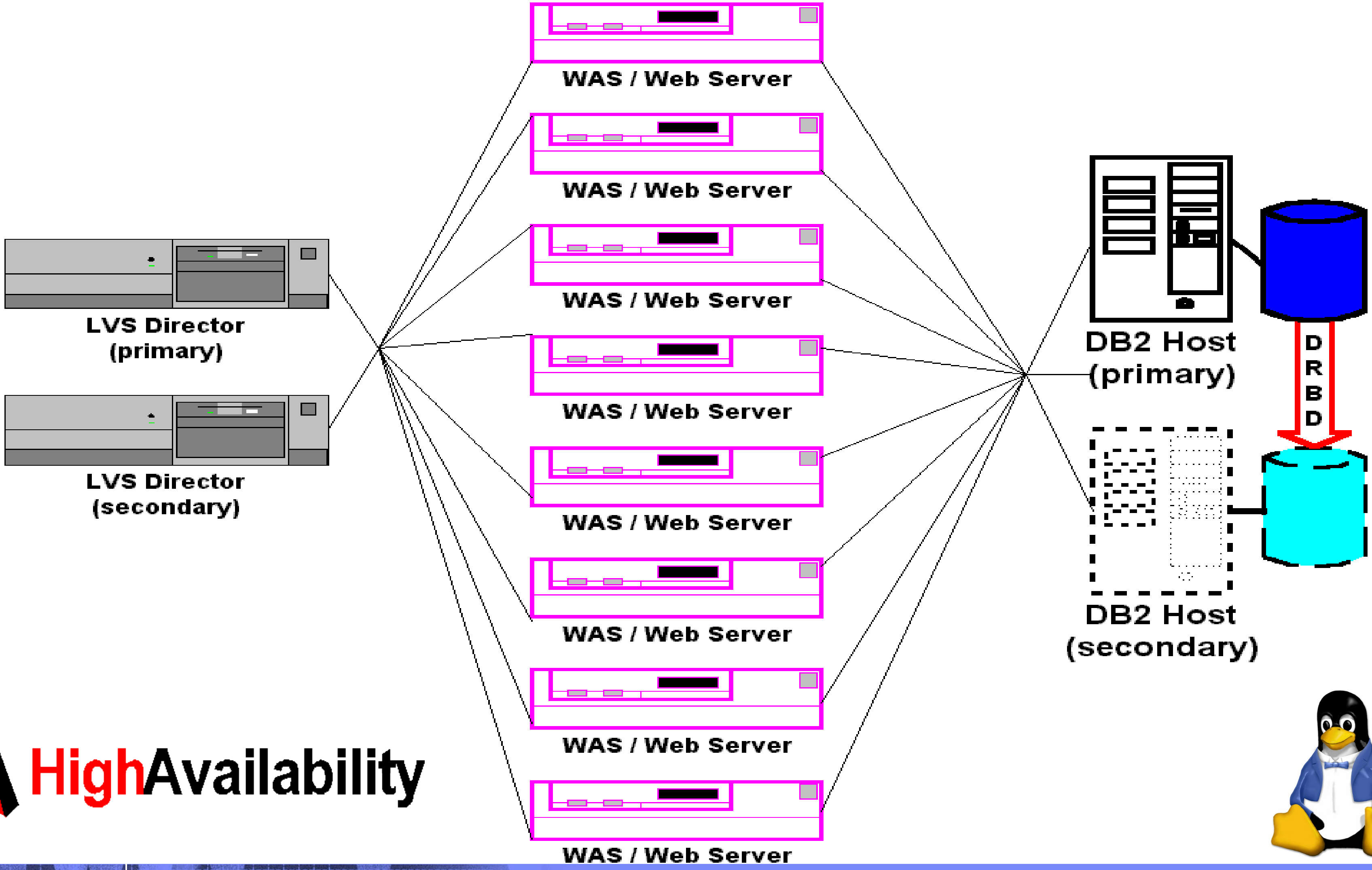
© 2006 IBM Corporation

Plays Well With Others

- Each of these independent services can work together to scale to large systems
- All single points of failure can be eliminated
- High-Availability, Load Balancing work together nicely



Linux-HA, DRBD and LVS Working Together



HighAvailability

© 2006 IBM Corporation

References

- http://linux-ha.org/
- http://www.drbd.org/
- http://www.linuxvirtualserver.org/

Legal Statements

- IBM is a trademark of International Business Machines Corporation.
- Linux is a registered trademark of Linus Torvalds.
- Other company, product, and service names may be trademarks or service marks of others.
- This work represents the views of the author and does not necessarily reflect the views of the IBM Corporation.

© 2006 IBM Corporation